The Plant Genome
Original Research

The FAD2 Gene Family of Soybean:

Insights into the Structural and Functional Divergence of a Paleopolyploid Genome

Jessica A. Schlueter, Iryna F. Vasylenko-Sanders, Shweta Deshpande, Jing Yi, Majesta Siegfried, Bruce A. Roe, Shannon D. Schlueter, Brian E. Scheffler, and Randy C. Shoemaker*

J.A. Schlueter and R.C. Shoemaker, USDA-ARS-CICGR, Ames, IA 50011; I.F. Vasylenko-Sanders, S. Deshpande, J. Yi, M. Seigfried, and B. Roe, Dep. of Chemistry and Biochemistry, Univ. of Oklahoma, Norman, OK 73019; S.D. Schlueter, Dep. of Genetics Development and Cellular Biology, Iowa State Univ., Ames, IA 50011; B. Scheffler, USDA-ARS MSA Genomics Lab., Stoneville, MS 38776. Received 13 June 2006. *Corresponding author (rcsshoe@iastate.edu).

Abstract

The ω-6 fatty acid desaturase (FAD2) gene family in soybean [*Glycine max* (L.) Merr.] consists of at least five members in four regions of the genome and are responsible for the conversion of oleic acid to linoleic acid. Here we report the identification of two new ω-6 fatty acid desaturase (FAD2) gene copies from soybean expressed sequence tags (ESTs). Four bacterial artificial chromosomes (BACs) containing five FAD2 genes were sequenced to investigate structural and functional conservation between duplicate loci. Sequence comparisons show that the soybean genome is a mosaic, with some duplicate regions retaining high sequence conservation in both genic and intergenic regions, while others have only the FAD2 genes in common. Genetic mapping using SSRs from within the BAC sequences showed that two BACs with high sequence homeology mapped to linkage groups I and O; these groups share syntenic markers. Another BAC mapped to linkage group L. The fourth BAC could not be mapped. Reverse transcriptase–polymerase chain reaction (RT–PCR) analysis of the five FAD2 genes showed that the FAD2-2B and FAD2-2C copies were the best candidates for temperature-dependent expression changes in developing pod tissue. Semiquantitative RT-PCR confirmed these results, with FAD2-2C showing upward of an eightfold increase in expression in developing pods grown in cooler conditions relative to those grown in warm conditions. The implications of these results suggest a candidate gene for controlling the levels of linoleic acid in developing pods grown in cooler climates.

Published in Crop Sci 47(S1) (2007). Published 14 Dec. 2006.
doi:10.2135/cropsci2006.06.0382tpg

677 S. Segoe Rd., Madison, WI 53711 USA

Cultivated soybeans provide greater than two thirds of the edible oil consumed in the USA, and 80% worldwide (Mattson et al., 2004). Soybean seed oil generally consists of varying amounts of oleic (12.2–22.4%) and linoleic acid (37.6–60.2%) depending on the cultivar.

Traditionally, oils for human consumption are preferred to have lower levels of polyunsaturated fatty acids, such as linoleic acid, that lead to autooxidation and undesirable flavor changes (Warner et al., 1994). Conversely, vegetable oils that are naturally higher in oleic acid are prized for having greater shelf-stability and increased structural integrity at higher cooking temperatures (Warner et al., 1994). They are also desirable because of the potential nutritional benefits, such as lowering low-density lipoprotein (LDL) cholesterol (Mattson and Grundy, 1985; Grundy, 1986).

Most polyunsaturated fatty acids, up to 90% in nonphotosynthetic tissues, are synthesized through

Abbreviations: BAC, bacterial artificial chromosome; DAE, days after emergence; DAF, days after flowering; EST, expressed sequence tag; gmw1, Williams 82 *G. max* BAC library constructed at Iowa State University; LTR, long terminal repeat; MYA, million years ago; PCR, polymerase chain reaction; qRT, semiquantitative reverse-transcriptase; RT, reverse transcriptase; SSR, simple sequence repeat.

an 18:1 desaturase in the endoplasmic reticulum (Miquel and Browse, 1994). The gene encoding microsomal ω-6 desaturase, fatty acid desaturase-2 (FAD2) was originally characterized in *Arabidopsis thaliana* (L.) Heynh. and associated with the conversion of 18:1 oleic to 18:2 linoleic acid by inserting a double bond at the 12th carbon in the fatty acid hydrocarbon chain (Okuley et al., 1994).

In soybean, two copies of microsomal ω-6 desaturase, FAD2-1 and FAD2-2, have been cloned (Heppard et al., 1996). Expression studies showed FAD2-1 seemed to be expressed primarily in developing seeds, while FAD2-2 was expressed in both vegetative tissues and developing seeds. Recently, EST based searches by Tang et al. (2005) identified an additional copy of FAD2-1 in soybean. An analysis of the EST libraries from which the FAD2 genes were identified found that all FAD2-1 ESTs were from seed-related libraries and FAD2-2 ESTs came from a variety of tissues.

A temperature-dependent relationship between oleic and linoleic acid concentration in soybean has been observed. Lower temperatures lead to an increase in polyunsaturated fatty acids, such as linoleic acid, and a decrease in unsaturated and monounsaturated fatty acids, such as oleic acid (Neidleman, 1987; Thompson, 1993; Thomas et al., 2003; Tang et al., 2005). The transcript levels of FAD2-1 and FAD2-2 did not appear to increase at low temperatures in developing pods as would be expected if they were responsible for the increased levels of linoleic acid (Heppard et al., 1996).

The existence of multiple FAD2 gene copies in soybean is not surprising, given that soybean is a paleopolyploid. Evidence for ancient polyploidy has been based on cytogenetic studies (Hadley and Hymowitz, 1973; Lackey, 1980; Walling et al., 2006), soybean gene family studies (Lee and Verma, 1984; Hightower and Meagher, 1985; Grandbastien et al., 1986; Nielsen et al., 1989), genetic mapping studies (Shoemaker et al., 1996; Lee et al., 1999, 2001), BAC hybridization, BAC-end sequencing and BAC fingerprinting (Marek et al., 2001; Foster-Hartnett et al., 2002; Yan et al., 2003, 2004), and EST based analyses that identified at least two major genome duplications (Schlueter et al., 2004; Blanc and Wolfe, 2004). Most recently, the first study looking at the sequence conservation in homeologous regions observed strong conservation of both gene order and orientation, as well as sequence conservation in the noncoding regions (Schlueter et al., 2006).

The FAD2 genes provide an excellent resource to further study the evolutionary dynamics of a paleopolyploid genome. In this paper, we report the identification of two additional FAD2 gene copies from soybean ESTs. Four soybean BACs representing five FAD2 gene copies were sequenced. Surprisingly, only two BACs showed high levels of homeology, while the other two BACs had only the FAD2-2 genes in common. The spatial and temporal expressions of each of the FAD2 genes were studied in developing pods grown at both warm and cool temperatures, as well as other vegetative tissues.

Materials and Methods

FAD2 BAC Selection

FAD2 genes were identified with TBLASTX (default parameters) using the FAD2-1A sequence (Genbank L43920) against all soybean ESTs (Altschul et al., 1990). Identified ESTs were aligned into contigs using Sequencher v. 4.5 (Gene Codes Corp., Ann Arbor, MI), also with default parameters. Four FAD2 copies were identified. The PCR primers were designed to distinguish between copies using Oligo 6.82 (Molecular Biology Insights, Cascade, CO); these sequences are in the supplemental materials. Multidimensional pools of the 'Williams 82' *G. max* BAC library (gmw1; Marek and Shoemaker, 1997) were PCR screened using the conditions described in supplemental materials. The BAC DNA was isolated using a Plasmid Midi kit (Qiagen, Valencia, CA) and reverified with PCR.

BAC Sequencing and Assembly

Three BACs, gmw1-45m6 (AC166742), gmw1-15k6 (AC160454), and gmw1-11j16 (AC166091) were sequenced at the University of Oklahoma. The detailed procedures for cloned large insert genomic DNA isolation, random shot-gun cloning, fluorescent-based DNA sequencing, and subsequent assembly were used as described previously (Bodenteich et al., 1993; Pan et al., 1994; Roe et al., 1996; Chissoe et al., 1995; Roe, 2004). The BAC DNA for gmw1-105h23 (AC187294) was randomly sheared, shot-gun cloned, sequenced, and assembled as previously described (Schlueter et al., 2006). Further details of these procedures can be found in the supplemental materials. The BAC sequences are available at NCBI under the above accession numbers.

Genetic Mapping of BACs

The BAC sequences were scanned for simple sequence repeats (SSRs) using Sputnik (Espresso Software Development, Seattle, WA). The SSR polymorphisms were identified in the *G. max* A81-356022 × *G. soja* PI 468916 mapping population (Diers et al., 1992; Shoemaker et al., 1996). Genetic map positions of these SSRs were determined using MapMaker with a minimum lod score of 3.0 (Diers et al., 1992; Lander et al., 1987). The relative positions of the BACs were then placed on the soybean composite map (Song et al., 2004; http://soybase.org, verified 22 Aug. 2006).

Sequence Annotation and Analysis

Gene structure prediction used a combination of ab initio and similarity-based methods displayed by the xGDB system (Schlueter et al., 2005; S.D. Schlueter, M.D. Wilkerson, Q. Dong, and V. Brendel, 2006, unpublished data). For ab initio prediction, Genscan with *A. thaliana* based parameters (Burge and Karlin, 1997), FgeneSH with *Medicago truncatula* Gaertn. based parameters (www.softberry.com; verified 13 Dec. 2006), and GeneMark.hmm with *A. thaliana* based parameters (Lukashin and Borodovsky, 1998) were run. For each BAC sequence, soybean ESTs and all plant putatively unique transcripts were aligned using Geneseqer at PlantGDB (Schlueter et al., 2003; Dong et al., 2005). For each gene, the structure was predicted using EST alignments, putatively unique transcript alignments, and then ab initio predictions with the yrGATE system as part of xGDB (Schlueter et al., 2005; Wilkerson et al., 2006). The xGDB database system in conjunction with yrGATE allowed the direct comparison of ab initio predictions to the EST alignments to better determine gene structures. Each predicted gene was subjected to a BLASTP query of the NCBI nr database, with default parameters to assign putative function as well as identify conserved domains (Altschul et al., 1990).

> The FAD2 genes provide an excellent resource to further study the evolutionary dynamics of a paleopolyploid genome.

Putative retroelements were identified with a variety of methods. Ab initio gene predictions identified some open reading frames that had sequence similarity to polyprotein sequences. The BLASTN and TBLASTX searches were performed against the TIGR repeat databases (www.tigr.org/tdb/e2k1/plant.repeats; verified 13 Dec. 2006), with default parameters. RepeatMasker was run with Repbase (A.F.A. Smit and P. Green, 2006, unpublished data), with default parameters. Potential long terminal repeat (LTR) retrotransposons were searched for using LTR_STRUC, with an intensity parameter of 1 (McCarthy and McDonald, 2003). Each BAC was fragmented into 1-kb pieces and a self-BLASTN, with default parameters, identified putative LTRs (Altschul et al., 1990).

Alignment of homeologous BACs used shuffle-LAGAN (Brudno et al., 2003), with default parameters anchored by predicted gene structures. Shuffle-LAGAN is a global pairwise alignment program that also detects rearrangements such as inversions. This produced a VISTA plot (Frazer et al., 2004). The nucleotide and protein percentage identity and similarity between genes both within a single BAC and between BACs and was calculated using WATER, a pairwise alignment program (gap penalty of 10; extension penalty of 0.2; EMBOSS). Synonymous and nonsynonymous distances were calculated using PAML with default parameters (Yang, 1997). Coalesence estimates were calculated as in Schlueter et al. (2004).

Transcript Accumulation of Homeologs

RT-PCR primers for all FAD2 genes as well as alternatively spliced transcripts were designed. Where possible, primers flanked an intron as an internal control and were designed to be homeolog specific. Each primer pair was tested by PCR against Williams 82 genomic DNA to verify the homeolog specificity.

Williams 82 developing pod tissue was grown in growth chambers with day/night temperature cycles of 32/28 or 18/12°C and light/dark cycles of 12/12 h, and was collected at 6 to 10 (5–10 mg), 13 (11–30 mg), 17 (31–60 mg), 19 (61–100 mg), 21 (101–150 mg), and 26 (200–300 mg) days after flowering (DAF) in both conditions as previously done by Heppard et al. (1996). Greenhouse grown tissue for cotyledons, roots, and furled unifoliate were collected 3 days after emergence (DAE), unfurled unifoliates at 4 DAE, cotyledons and roots at 7 and 8 DAE, respectively, furled trifoliolate at 11 DAE, unfurled trifoliolate at 15 DAE, and flower tissue ≈60 DAE. For each tissue, samples were collected from at least three independent plants. Using the RNeasy Plant Mini Kit (Qiagen, Valencia, CA), mRNA was extracted and purified from flash-frozen tissue and subsequently treated with DNA-free DNase and removal kit (Ambion, Austin, TX) and quantified using a ND-1000 spectrophotometer (NanoDrop, Wilmington, DE).

Nonquantitative RT-PCR was performed with SuperScript One-Step RT-PCR (Invitrogen, Carlsbad, CA) across all tissues. The PCR conditions were as previously described (Schlueter et al., 2006) and are provided in the supplemental materials. Tubulin was used as a positive control (Graham et al., 2002). All RT-PCR reactions were done with three independent biological replicates.

Semiquantitative real-time PCR was performed with all FAD2 genes (and alternatively spliced transcripts) that showed expression with the initial RT-PCR screens using only the developing pod tissues. Reactions were done using Stratagene's Brilliant qRT–PCR kit (La Jolla, CA) and the Stratagene

Mx3000P thermocycler. Each sample contained a passive reference dye and was run in triple technical replicates. Each primer pair was additionally run on two biological replicates. Reaction conditions are described in the supplemental materials.

Results

Genomic Organization of FAD2 Regions

The PCR-based screens of the gmw1 BAC library identified four BACs each for FAD2-1A, FAD2-1B, FAD2-2A, and FAD2-2C. To maximize regional overlap, the largest BAC corresponding to each gene was sequenced. Shotgun sequencing of BACs gmw1-15k6 (FAD2-1A), gmw1-105h23 (FAD2-1B), gmw1-11j6 (FAD2-2A), and gmw1-45m6 (FAD2-2C) yielded >497 kb of soybean genomic sequence anchored by FAD2 genes. The SSRs identified from each BAC allowed genetic mapping of these BACs to position gmw1-15k6 and gmw1-105h23 on linkage groups O and I, respectively (Fig. 1). The BAC gmw1-11j16 mapped to linkage group L, at a position 24.3 cM from peG488_1 and 23.0 cM from Satt156 (results not shown). There were no close associations between fatty acid or oil QTLs and the determined map locations. Ten different SSRs were tested from gmw1-45m6, and none identified polymorphisms in the mapping population screened.

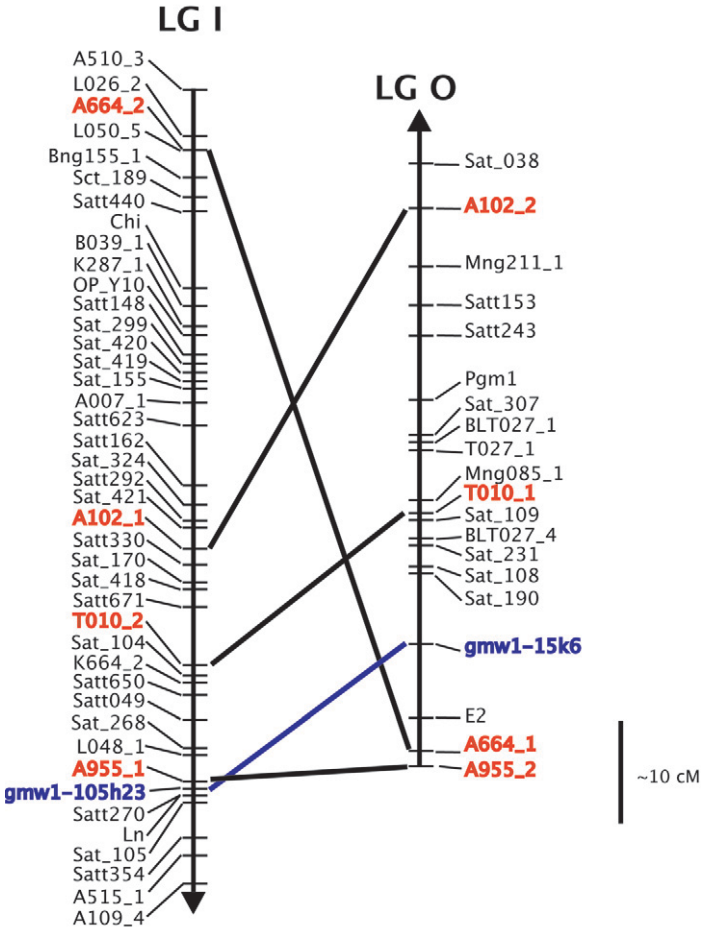

Figure 1. Graphical representation of Linkage Groups I and O based on the soybean composite map at www.soybase.org (verified 13 Dec. 2006). The BACs gmw1-15k6 (FAD2-1A) and gmw1-105h23 (FAD2-1B) mapped to these linkage groups in the *Glycine max* A81-356022 × *Glycine soja* PI 468.916 mapping populations. Their locations relative to other markers are shown. Lines between markers show syntenic markers.

Gene predictions for all four BACs identified a total of 57 genes (Fig. 2, Supplemental Table 1). All of these gene structures, EST alignments, and ab inito predictions are available to view at http://soybase.org/publication_data/Schlueter/GMaxGDB.html (verified 13 Dec. 2006). Of all predicted genes, 65.5% possessed exons that were supported by EST alignments, similar to an estimate from a previous study (65%; Schlueter et al., 2006). The average gene size was ≈956 bp. The gene density (not including repetitive elements) in these regions seems similar in three BACs with 1 gene every 6.70, 7.95, and 7.77 kb for gmw1-15k6, gmw1-105h23, and gmw1-11j16, respectively. Conversely, gmw1-45m6 has only 1 gene every 19.2 kb. These predictions, not considering gmw1-45m6, fall in the range of previous gene density estimates ranging from 1 gene per 5.8 to 9.9 kb (Young et al., 2003; Mudge et al., 2005).

All of the translated protein sequences of predicted genes showed similarity to a protein in the NCBI nonredundant database (Supplemental Table 1). On the basis of the EST alignments, FAD2-1A and FAD2-1B both show evidence of alternatively spliced transcripts. For both FAD2-1A (gmw1-15k6) and FAD2-1B (gmw1-105h23), the alternative spliced transcript moves the start methionine codon by increasing (or decreasing) the intron size. This alternative splicing event does not, however, change the resulting protein. These transcripts are called FAD2-1A_S and FAD2-1A_L for the alternative splicing events that have a small intron (S) or a large intron (L), respectively. The same terminology applies to FAD2-1B. Other potential alternative splicing events include a vesicle-associated membrane protein, a ribonuclease HII protein, and a protein most like a *Cicer arietinum* L. gene (CAD31715.1) on gmw1-15k6. On gmw1-105h23, melvonate dispohosphate decarboxylase and a pollen-specific protein show alternative splicing (results shown at http://soybase.org/publication_data/Schlueter/GMaxGDB.html; verified 13 Dec. 2006)). In contrast to evidence of alternative

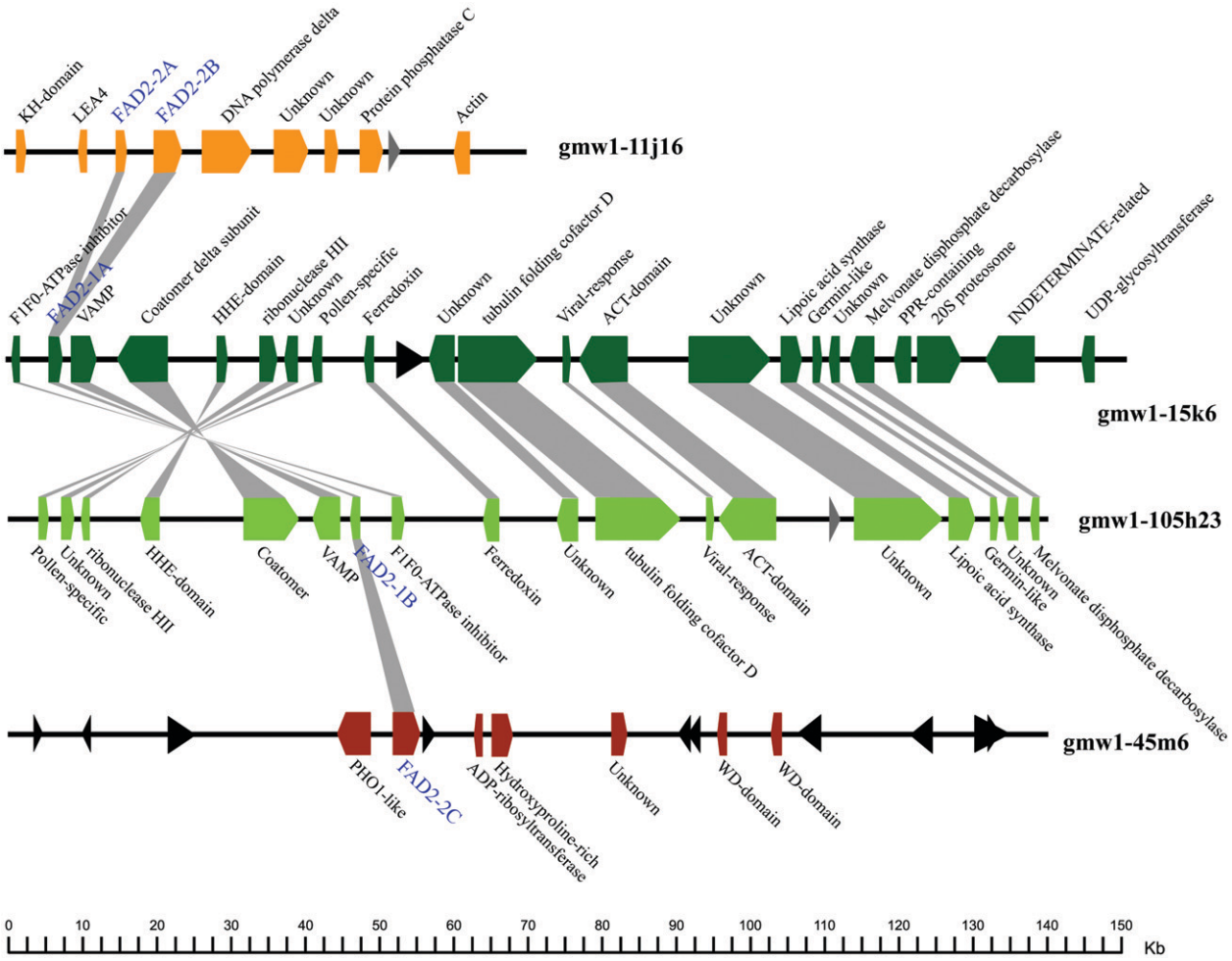

Figure 2. Predicted gene structures and retrotransposon insertions on soybean BACs gmw1-11j16, gmw1-15k6, gmw1-105h23, and gmw1-45m6. Each colored block arrow on the line represents a gene. Gray connecting boxes between genes show homeologs or syntenic relationships. Black arrowheads represent full-length predicted retrotransposon elements. Gray arrowheads represent fragmented remnants of retrotransposons.

splicing, there appeared to be only two cases of tandem duplication, the WD-domain containing genes on gmw1-45m6 and the FAD2-2 genes (A and B) on gmw1-11j16 (Fig. 2).

Only sequences with fragmented similarity to retroelement RTs were identified in gmw1-11j16 and gmw1-105h23. The putative retroelement sequence in gmw1-11j16 is similar to a gag-pol polyprotein previously identified in *G. max* (AAC64917.1), although the BLAST search showed only a partial alignment. Gmw1-105h23 has a degenerate *Ty1-copia* like RT at ≈112 250–113 000 bp. Gmw1-15k6 contains only one retrotransposon inserted between the ferredoxin gene and an unknown protein similar to *Oryza sativa* NP_915582.1. This retroelement is nearly identical to a previously identified gypsy-like retroelement in soybean (AAO23078.1), with an e value (BLAST statistic, measure of similarity between two sequences) of 0.0 (highly significant, near 100% sequence identity) and 1379/1554 positives. The gypsy element structure with a RT preceding the integrase is conserved in this newly identified retrotransposon. The LTRs for this element are only 363 bp in length, but are highly similar to one another, suggesting a recent insertion.

The BAC gmw1-45m6 is very different from the other three BACs due to the presence of numerous repetitive element sequences. A BLASTX search of known RT sequences against the BAC identified seven putative Diaspora *Ty3-gypsy* like elements. The program LTR_STRUC identified one LTR-retrotransposon of >15706 bp in length. When searching for the location of the LTRs for the LTR_STRUC predicted element, an interesting trend occurred: eleven different regions of gmw1-45m6 showed high sequence identity to that LTR sequence. This particular LTR-retrotransposon has inserted itself numerous times across this region. Additionally, a putative CACTA element fragment was identified from 57 875–59 365

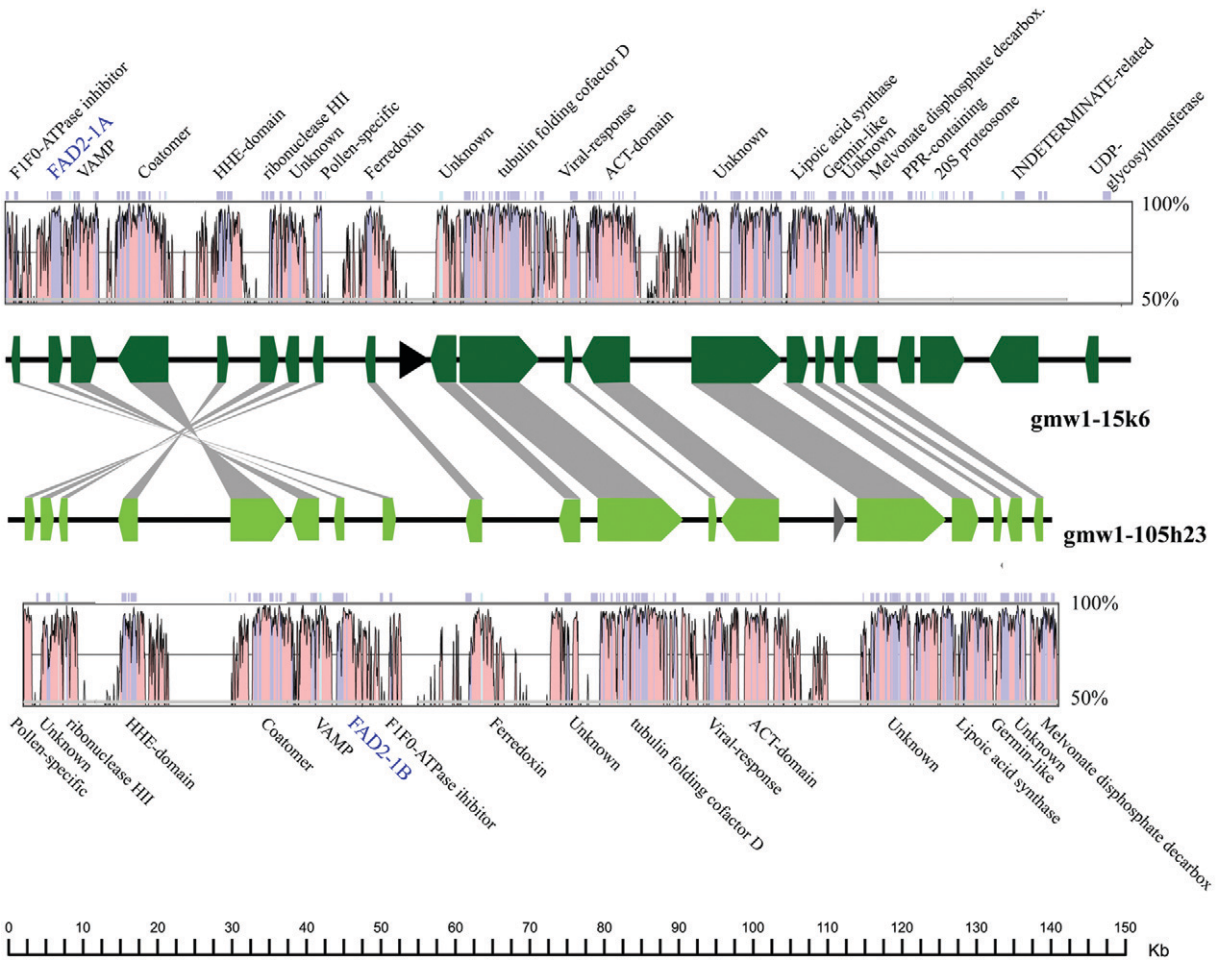

Figure 3. VISTA based identity plots between gmw1-15k6 and gmw1-105h23. The plots above and below the representative BAC structures show the relative nucleotide identity between the two BACs. The light purple boxes on top of the VISTA plots correspond to exon positions.

bp (just 3′ of the FAD2-2C gene copy) based on a TBLASTX similarity search to an *A. thaliana* CACTA1, with an e value of 0.0. However, only a small region of the CACTA element was conserved.

Comparison of FAD2 Homeologous Regions

Out of the four FAD2-containing BAC clones, only two of them share several other genic sequences (Fig. 2). The homeology between gmw1-105h23 (FAD2-1A) and gmw1-15k6 (FAD2-1B) is high (Fig. 3). One major inversion spanning the region from an F1F0-ATPase inhibitor-like gene to the pollen-specific gene is observed. With this exception, all of the genes are conserved in both order and orientation in these regions. Some sequence similarity (>75%) is seen in intergenic regions (Fig. 3). Because this is seen in close proximity to the coding regions, it is likely due to conserved promoter elements and transcription factor binding sites. There are, however, large regions between genes that have <50% sequence identity; in other words, they are not conserved at all.

Similar to previous findings (Schlueter et al., 2006), the homeolog gene length is preserved. Between each gene pair, the number of exons and introns for most genes are identical. Exceptions to this are as follows: the gmw1-105h23 ribonuclease HII gene that appears to be truncated with respect to the gmw1-15k6 gene; the gmw1-15k6 F1F0-ATPase inhibitor-like gene appears to have an included intron that is not found in the gmw1-105h23 copy; four exons in the gmw1-105h23 unknown gene most like *Oryza sativa* NP_915582.1 are not predicted in the gmw1-15k6 copy; one exon on the gmw1-105h23 unknown gene most like *Arabidopsis* NP_197637.1 is not predicted in the gmw1-15k6 copy; and the alternative splicing events as discussed above. The gmw1-15k6 copy of melvonate diphosphate decarboxylase contains five more exons relative to the gmw1-105h23 copy, although this is most likely due to truncation of the gmw1-105h23 copy since it is

Table 1. Similarity between homeologous BACs gmw1-15k6 and gmw1-105h23.

Putative function	Nucleotide identity	Protein identity	Protein similarity	K_s†	K_a‡	Date (MYA)§
Unknown 5	88.4	81.6	86.2	0.1280	0.0880	10.5
FAD2-1	94.6	93.5	96.9	0.1048	0.0329	8.59
Vesicle-associated membrane protein	97.3	97.9	98.7	0.0749	0.0116	6.14
Coatomer	97.2	97.3	98.3	0.0611	0.0131	5.01
HHE-domain containing	93.4	93.4	95.2	0.1360	0.0295	11.1
Ribonuclease HII	88.2	79.7	81.4	0.0268	0.0557	2.20
Unknown 70	89.4	90.4	91.5	0.2361	0.0257	19.4
Pollen-specific	93.5	91.2	93.4	0.1065	0.0514	8.73
Ferredoxin	94.8	94.8	98.7	0.1345	0.0247	11.0
Unknown 110	90.1	84.3	82.2	0.0735	0.0337	6.02
B-tubulin cofactor D	91.6	88.5	86.8	0.1012	0.0215	8.30
Viral-response protein	93.0	91.3	93.6	0.1007	0.0275	8.25
ACT-domain protein	97.4	97.5	98.6	0.0557	0.0115	4.57
Unknown 130	69.7	66.8	69.0	0.1020	0.0533	8.36
Lipoic acid synthase	91.9	90.9	92.3	0.1129	0.0260	9.25
Germin-like protein	95.9	96.6	97.8	0.1207	0.0156	9.89
Unknown 160	69.8	66.8	68.9	0.1328	0.0123	10.9
Melvonate diphosphate decarboxylase	96.1	97.2	97.2	0.1013	0.0532	8.30
Average	90.7	88.9	90.4	0.1061	0.0326	8.70

† K_s, synonymous distance between retained homeologs.
‡ K_a, nonsynonymous distance between retained homeologs.
§ MYA, million years ago.

Table 2. FAD2 primers for semiquantitative RT-PCR screens.

Primer name†	Primer sequence (5′–3′)	Primer length bp	Primer location‡	Size§
gmw1-15k6 FAD2-1A_L F	GTT CCC AAG AGT ATA AAA CTG C	22	5553–5574	843
gmw1-15k6 FAD2-1A_L R	CAC AAC ATC ATC AAC CCA TTG	21	6376–6396	
gmw1-15k6 FAD2-1A_S F	GAT TTC CTG AAG GCT TAG GTG	21	5717–5737	452
gmw1-15k6 FAD2-1A_S R	GAA AGC AGT GTG GTG GAA TTG	21	6149–6169	
gmw1-105h23 FAD2-1B_L R	GAT ACT ACA AGC CAC TAG GCA T	22	91152–91173	821
gmw1-105h23 FAD2-1B_L R	TGA GTG AAC GGT CAA ACC CAT A	22	91952–91973	
gmw1-105h23 FAD2-1B_S F	GAT TTC TTG AAG GCT TAT GGT AT	23	91292–91314	480
gmw1-105h23 FAD2-1B_S R	AAG GTC ATA AAC AAC ATA GGA C	22	91751–91772	
gmw1-11j16 FAD2-2A F	CTT CCA CAA CCT TCC TCA TC	20	14964–14713	335
gmw1-11j16 FAD2-2A R	GGA TTG TTT AGG TAT TTA GAT AG	23	15277–15299	
gmw1-11j16 FAD2-2B F	AGT ATG CTA TGG CCT TTT CTG	21	23801–23821	494
gmw1-11j16 FAD2-2B R	ACC ATC CCA CCT GAC GGA C	19	24277–24295	
gmw1-45m6 FAD2-2C F	ACC TCA CCA TAG CCT TCT AC	20	54676–54695	336
gmw1-45m6 FAD2-2C R	TAT TTA GAG TAC CAC TTG ATA CA	23	54990–55012	

† F corresponds to the forward primer and R corresponds to the reverse primer.
‡ Location of primer within BAC sequence, relative to 5′ end of BAC.
§ Size of PCR amplicon.

at the end of the BAC. The average nucleotide identity between the coding regions of gmw1-15k6 and gmw1-105h23 is 90.7%; and the resulting amino acid identity and similarity is 88.9 and 90.4%, respectively (Table 1).

Synonymous and nonsynonymous distance measures between homeologs on gmw1-15k6 and gmw1-105h23 were calculated. Two independent studies using EST sequences and different methods have predicted that the soybean genome underwent major duplications ≈14.5 and 45 million years ago (MYA; Schlueter et al., 2004) or ≈4.5 and 16.1 MYA (Blanc and Wolfe, 2004). The predicted coalescence estimate, based on the average synonymous distance of homeologs between gmw1-15k6 and gmw1-105h23 (0.1061), is ≈8.70 MYA (Table 2). A *t* test using the mean of synonymous distances from the FAD2 regions and the mean of synonymous distances from the 14.5 MYA event (Schlueter et al., 2004) gave a *t* value of 7.57 with 17 df, showing a significant difference between the mean synonymous distances with a >99% confidence.

The FAD2 Gene Family

When only the FAD2 genes are considered, some interesting trends emerge. On average, the nucleotide identity between the FAD2 gene copies is 76.48% with a range of 66.6 to 95.8%. The average protein identity is 77.94% with a range of 67.4 to 95.9%, and the average protein similarity is 85.12% with a range of 72.1 to 96.9%. Naturally, FAD2-1A (gmw1-15k6) and FAD2-1B (gmw1-105h23) are most similar to one another. The overlapping regions of FAD2-2B (gmw1-11j16) and FAD2-2C (gmw1-45m6) have high nucleotide identity (95.8%), as well as high protein identity (95.9%) and similarity (96.6%). There has been surprising conservation in these FAD2-2 coding regions with seemingly nothing else in common between the BACs.

Expression of FAD2 in Developing Pod Tissue

Of particular agronomic interest is the functional divergence between the FAD2 genes, especially in

developing pod tissues grown at warm or cool temperatures. Reverse-transcriptase PCR screens were performed across developing pod tissues as well as vegetative tissues with all FAD2 genes and their alternatively spliced transcripts to identify tissues where each transcript accumulated (Fig. 4). The FAD2-2A (gmw1-11j16) primer pairs did not amplify any RNA samples, although the positive Williams 82 genomic DNA control amplified (data not shown). This suggests that FAD2-2A is either not expressed in tissue surveyed in this experiment, or that this gene copy is no longer functional. Detection of all the alternatively spliced versions of FAD2-1A and FAD2-1B transcripts verified our prediction of alternatively spliced structures.

The RT-PCR screens of all four FAD2-1A (gmw1-15k6) and FAD2-1B (gmw1-105h23) transcripts suggest that none have increased expression in developing pods at cooler conditions. These results confirm previous findings (Heppard et al., 1996). In fact, there appears to be delayed or less expression of FAD2-1A and FAD2-1B in developing pods grown in cooler conditions. What was surprising was the amplification of both FAD2-1A_L (gmw1-15k6) and FAD2-1B_S (gmw1-105h23) in vegetative tissues when these copies were originally shown to express only in developing pods (Fig. 4; Heppard et al., 1996). In concordance with previous results (Heppard et al., 1996), FAD2-2B (gmw1-11j16) is expressed in all tissues surveyed but the intensity of the bands in developing pods grown at cooler conditions suggests that FAD2-2B may be a candidate for increased expression of FAD2 (Fig. 4). FAD2-2C (gmw1-45m6) shows very similar expression patterns to FAD2-2B.

Following the results of the traditional RT-PCR screens, semiquantitative reverse-transcriptase (qRT) PCR experiments were performed. The objective was to determine if significant fold changes in expression could be identified at each time point for developing pods grown under cool conditions vs. those grown under warm conditions. The number of cycles to reach threshold fluorescence (the Ct value) was determined based on the point at which the FAD2 amplicon fluoresced at a higher level than

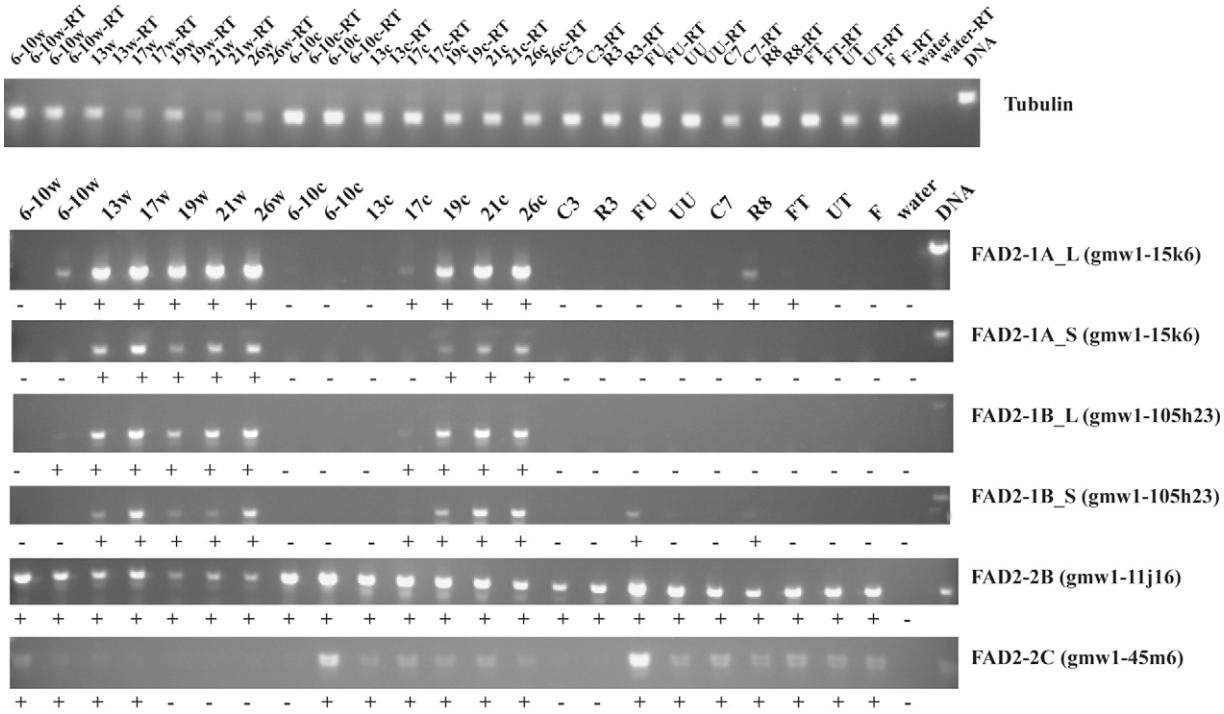

Figure 4. RT-PCR amplification for FAD2-1A_L, FAD2-1A_S, FAD2-1B_L, FAD2-1B_S, FAD2-2B, and FAD2-2C. Positive control reactions used tubulin56. Plus symbols indicate amplification of that transcript in a particular tissue, and minus symbols indicate no amplification. Tissue types are as follows: 6–10, developing pods 6–10 d after flowering (DAF); 13, developing pods 13 DAF; 17, developing pods 17 DAF; 19, developing pods 19 DAF; 21, developing pods 21 DAF; 26, developing pods 26 DAF; C3, cotyledons 3 days after emergence (DAE); R3, roots 3 DAE; FU, furled unifoliate 3 DAE; UU, unfurled unifoliate 4 DAE; C7, cotyledons 7 DAE; R8, roots 8 DAE; FT, furled trifoliolate 11 DAE; UT, unfurled trifoliolate 15 DAE; P, pods 76 DAE. All samples with a w were grown under the warm conditions and c for cool conditions. All samples with a -RT are controls with no reverse transcriptase for mRNA amplification, but still contain Taq DNA polymerase. The final sample in each set of reactions is a Williams 82 DNA based positive control.

Table 3. Semiquantitative reverse-transcriptase PCR results.

Gene	Tissue DAF¶	Temperature	Ct value†	Change in Ct value‡	Fold change in expression§
FAD2-1A	6–10	warm	35.5	0.9	
		cool	34.6		
	6–10	warm	34.5	–0.2	
		cool	34.7		
	13	warm	28.9	–6.2	–12.4
		cool	35.1		
	17	warm	28.1	–7.1	–14.2
		cool	35.2		
	19	warm	30.1	–0.2	
		cool	30.3		
	21	warm	28.3	–0.6	
		cool	28.9		
	26	warm	27.4	–2.9	–5.8
		cool	30.3		
FAD2-1B	6–10	warm	39.0	0.9	
		cool	38.1		
	6–10	warm	38.8	0	
		cool	38.8		
	13	warm	33.9	–4.2	–8.4
		cool	38.1		
	17	warm	32.5	–5.6	–11.2
		cool	38.1		
	19	warm	35.7	1.1	2.2
		cool	34.6		
	21	warm	34.5	0.8	
		cool	33.7		
	26	warm	33.2	–1.7	–3.4
		cool	34.9		
FAD2-2B	6–10	warm	32.3	–0.4	
		cool	32.7		
	6–10	warm	33.1	0.7	
		cool	32.4		
	13	warm	32.3	0.3	
		cool	32.0		
	17	warm	32.1	–0.2	
		cool	32.3		
	19	warm	33.1	0.9	
		cool	32.2		
	21	warm	32.7	0.4	
		cool	32.3		
	26	warm	32.5	–0.6	
		cool	33.1		
FAD2-2C	6–10	warm	33.4	1.3	2.6
		cool	32.1		
	6–10	warm	32.6	–1.0	–2.0
		cool	33.6		
	13	warm	34.6	2	4.0
		cool	32.6		
	17	warm	33.9	0.5	
		cool	33.4		
	19	warm	36.4	4	8.0
		cool	32.4		
	21	warm	35.6	2.7	5.4
		cool	32.9		
	26	warm	34.8	2.2	4.4
		cool	32.6		

† The Ct value is the threshold where product fluorescence levels are statistically higher than background fluorescence.
‡ Change in Ct value is relative to developing pods in warm conditions.
§ The fold change in expression is based on a twofold change relative to one degree in the Ct value.
¶ DAF, days after flowering.

background fluorescence. A significant expression difference between cool pod tissue and warm pod tissue at a particular time point corresponds to a change of one cycle in Ct value; a one-cycle change translates into a twofold change in expression. The qRT–PCR screens were run with the six FAD2 primer pairs that detected expression in developing pods (Fig. 4). Only the developing pod tissue was used in these experiments. Both FAD2-1A_L (gmw1-15k6) and FAD2-1B_L (gmw1-105h23) show significant changes in expression in developing pods from warm conditions vs. those from cool conditions (Table 3). In both cases, expression of FAD2 increases in the warm pod tissues relative to the cool pod tissue (Table 3). The qRT–PCR results for the alternative spliced FAD2-1A_S and FAD2-1B_S transcripts mirrored those of FAD2-1A_L and FAD2-1B_L (data not shown).

Although RT-PCR screens suggested that FAD2-2B (gmw1-11j16) seemed a good candidate for temperature-dependent increase in FAD2 expression, the qRT–PCR did not confirm this. FAD2-2C is the only ω-6 desaturase gene that shows increases in expression in the cool developing pods. Fairly significant fold changes in expression are observed with FAD2-2C (gmw1-45m6) at 6–10, 13, 19, 21, and 26 DAF ranging from 2.6 to 8.0 (Table 3).

Discussion

Paleopolyploid Genome Structure

Paleopolyploids are ancient polyploids that are undergoing the switch from tetrasomic to disomic inheritance in a process called diploidization. It has been previously suggested that during this ongoing process of diploidization there are both diploid and tetraploid loci within the genome (Wolfe, 2001). Most plant species are now considered to be paleopolyploids (Lockton and Gaut, 2005; Blanc and Wolfe, 2004; Schlueter et al., 2004; Masterson, 1994). The first study looking at homeologous regions in soybean suggested that duplicated regions have higher-than-expected gene retention and that, particularly in soybean, the process of diploidization is quite slow and certainly incomplete (Schlueter et al., 2006). Two BACs sequenced in this analysis support previous findings. However, the other

two BACs show very different evolutionary dynamics within the soybean genome.

Between the soybean composite maps for linkage groups I and O, there are numerous markers in common, suggesting that gmw1-105h23 and gmw1-15k6 are indeed homeologs (Fig. 1). The mapping of gmw1-11j16 to linkage group L is significant because this linkage group has been shown, through FISH analysis, to have secondary hybridization signals on another chromosome not yet anchored to a linkage group (Walling et al., 2006). This suggests that there may be another region in the genome that is homeologous to gmw1-11j6 and that there may be more copies of FAD2. Similarly, hybridization of a 40 bp overgo to the Gmw2 Williams 82 BAC library membranes identified at least 52 BACs. Given that these membranes represent ≈7.92× genome coverage, this suggests that there are ≈6.6 FAD2 loci in the genome. This translates to possibly two or more regions of the genome beyond those studied here that contain FAD2 genes.

Between gmw1-15k6 and gmw1-105h23, genic sequence is strongly retained in both order and orientation (Fig. 2). Gene structure variation between the BACs may be due to the lack of EST support for particular gene structures and the reliance on ab inito predictions. These two BACs have an even higher level of sequence retention with no tandem duplications or gene losses, as previously observed. The coalescence estimate between these BACs is 8.70 MYA (Table 1); a much more recent duplication event than the 12.2 MYA estimate obtained for the HCBT duplicated BACs (Schlueter et al., 2006). Further, this divergence estimate does not fall within the closest duplication estimate of 14.5 MYA as shown by *t* test (Schlueter et al., 2004). It is possible that this duplication is the result of a segmental or aneuploidy event that happened independently of the paleopolyploid events. Or, this region may be under selection forces that have conserved gene structures more than other regions of the genome.

FAD2-1A and FAD2-1B genes yielded a divergence estimate of ≈8.59 MYA, close to the average of all homeologs on those BACs. However, when these two closely related FAD2-1 copies are compared with the FAD2-2 genes, synonymous distances suggest that they diverged from 105 to 277 MYA. When just the FAD2-2 genes are compared with one another, it appears that the tandem duplication on gmw1-11j16 occurred about 28 MYA and that FAD2-2B is more closely related to the FAD2-2C (gmw1-45m6) with a coalescence estimate of 10 MYA. This is reflected as well in the sequence identity discussed above. Similarly, FAD2-2A and FAD2-2C distances suggest a divergence at 31.9 MYA. This estimate is relatively close to that of the tandem duplication on gmw1-11j16. One possibility is that FAD2-2B was tandemly duplicated, and then the region itself was duplicated (leading to the FAD2-2C BAC), followed by extensive loss, rearrangement, and/or retroelement insertion. Another possibility would be that FAD2-2C is the result of a single gene insertion into a region of the genome that is very repetitive, or that FAD2-2C or FAD2-2B are closely related due to gene conversion. Taken together, these results suggest that the duplication that generated the FAD2-1 and FAD2-2 genes is quite ancient.

Current plant genome structure models suggest that genes are arranged in two major ways. In smaller genomes such as *A. thaliana*, the genes are fairly evenly spaced along the chromosomes as "beads on a string" (The Arabidopsis Genome Initiative, 2000). Conversely, the maize genome has been predicted to have genes existing in islands separated by repetitive sequence (SanMiguel and Bennetzen, 1998; Messing et al., 2004), although this model has recently been called into question (Haberer et al., 2005). When all four BACs are considered, support for both genome models is observed. There are three cases of regions with fairly densely packed genes with almost no retroelement insertions (Fig. 2), and the last region has very few genes with large retroelement insertions and is nearly half as dense as the other FAD2 BACs. The genes in this region are clustered and surrounded by numerous repetitive sequences much as with the gene island model. Although there is little to no evidence for a transposon explosion in the soybean genome, organization of this BAC would suggest that some regions might be hotspots for retroelement insertions. These varying gene densities suggest that the soybean genome does not fit a single model for genome organization and appears to be a mosaic, at least in portions of the genome. Most of the gene space is probably composed of fairly densely packed genic regions while other parts of the genome have sparse genes interspersed with repetitive elements. Further, gene space may not be restricted to nonrepetitive regions as has previously been suggested (Lin et al., 2005).

Functional Divergence of FAD2 Genes

The FAD2 genes provide an excellent model for studying functional divergence in a paleopolyploid genome. Previous studies have identified FAD2 genes as encoding ω-6 fatty acid desaturases that catalyze the conversion of 18:1 oleic acid to 18:2 linoleic acid in soybean (Okuley et al., 1994; Heppard et al., 1996). This conversion is of particular interest in soybean due to the growing desire for low-linoleic soybean oil (Thomas et al., 2003). The structure of the five

sequenced FAD2 genes in soybean are similar to one another but form two distinct sub-groups, the FAD2-1's and FAD2-2's. Previous work has looked at the expression of FAD2-1A and FAD2-2B in both vegetative tissues and developing seeds (Heppard et al., 1996). Their results suggested that FAD2-1A is expressed specifically in seeds and that FAD2-2B is expressed in all tissues, but at lower levels in seeds than FAD2-1. This partitioning of function is very characteristic of possible fates of duplicate genes in a paleopolyploid. Following duplication, paralogs may either retain original gene function, subfunctionalize, neofunctionalize (obtain a new function), or be silenced (Force et al., 1999). Further, when looking for temperature-dependent expression differences with either FAD2 copy, none were observed (Heppard et al., 1996). This was surprising since developing pods grown at cooler temperatures have an increase in linoleic acid, presumably due to increased expression of a FAD2 gene.

The identification of three more copies of FAD2 raises several questions as to the expression of each gene: are both FAD2-1s only expressed in pod tissue, and is there a temperature-dependent increase in expression in any FAD2 gene? RT-PCR screens using FAD2 primers showed that FAD2-1 (gmw1-15k6) does not appear to be expressed exclusively in pods. While their expression is primarily in developing pod tissues, both FAD2-1A_L (gmw1-15k6) and FAD2-1B_S (gmw1-105h23) showed expression in other tissues. These tissues represent different developmental stages of young seedlings; FAD2-1B_S may be expressed earlier in the seedling followed by expression of FAD2-1A_L.

Alternative splicing is another means for functional divergence (Su et al., 2006). The EST-based gene structure prediction identified at least two functional cases of alternative splicing in FAD2 genes, FAD2-1A (gmw1-15k6) and FAD2-1B (gmw1-105h23). Looking at the similar structure of the alternative spliced products, it appears that the FAD2-1 genes retained their alternatively spliced transcripts after duplication. Conversely, FAD2-2 ESTs did not show evidence for alternative splicing. Instead, FAD2-2 has been tandemly duplicated on gmw1-11j16. This shows that even within a gene family, the mechanisms for increasing the material for functional divergence may be varied. The RT-PCR results show slight differences between the alternatively spliced transcripts, but only in the vegetative tissues (Fig. 4). Similarly, when semiquantitative RT-PCR reactions were performed, no differences could be identified in the developing pod tissue between the alternatively spliced products (data not shown). In both cases, when looking for changes in fold expression at each time point between developing pods from cool conditions and warm conditions, the fold change was the same for the alternatively spliced transcripts of FAD2-1A and FAD2-1B.

The primary objective of the semiquantitative RT-PCR experiments was to look for fold change differences in expression between developing pods from cool and warm conditions. Table 3 shows that FAD2-2C (gmw1-45m6) is the only gene that showed significant transcript increase in developing pods from cool conditions. Thus, FAD2-2C is the most likely gene candidate for increasing the pool of ω-6 fatty acid desaturases at lower temperatures leading to an increase in linoleic acid. This gene, however, is not the only one to show temperature-dependent changes in transcript accumulation. Both FAD2-1A (gmw1-15k6) and FAD2-1B (gmw1-105h23, and their respective alternative transcripts) show significant fold increases in transcript accumulation in developing pods from warm conditions. These results are similar to those shown with Northern blot analysis by Heppard et al. (1996). This increase in expression may be due to the proposed instability of the FAD2-1 encoded ω-6 fatty acid desaturases at higher temperatures (Heppard et al., 1996; Tang et al., 2005). Further studies looking at the expression of each of these genes in the various compartments of developing pods may identify varying expression patterns of the FAD2 genes or even a partitioning of expression. The proposed role of modifying genes in the conversion of oleic to linoleic acid (Alt et al., 2005) should also be considered.

Acknowledgments

The authors would like to thank Mary Duke and Xiaofen Liu for their sequencing expertise. We would also like to thank Jody Hayes for her skills in mapping SSR markers, Terry Olsen for his MapMaker expertise, and Greg Peiffer for his assistance in identifying and subcloning BACs. This article is a contribution of the Corn Insect and Crop Genetics Research Unit (USDA-ARS). J.A. Schlueter was supported by a grant from the United Soybean Board.

References

Alt, J.L., W.R. Fehr, G.A. Welke, and D. Sandhu. 2005. Pheotypic and molecular analysis of oleate content in the mutant soybean line M23. Crop Sci. 45:1997–2000.

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403–410.

The Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature (London) 408:796–815.

Blanc, G., and K.H. Wolfe. 2004. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell 16:1667–1678.

Bodenteich, A., S. Chissoe, Y.F. Wany, and B.A. Roe. 1993. Shotgun cloning as the stragety for choice to generate templates for high-throughput dideoxynucleotide sequencing. p. 42–50. *In* J.C. Venter (ed.) Automated DNA sequencing and analysis techniques. Academic Press, London.

Brudno, M., S. Malde, A. Poliakov, C.B. Do, O. Couronne, I. Dubchak, and S. Batzouglou. 2003. Glocal alignment: Finding rearrangements during alignment. Bioinformatics 19S1:i54–i62.

Burge, C., and S. Karlin. 1997. Prediction of complete gene structures in human genomic DNA. J. Mol. Biol. 268:78–94.

Chissoe, S.L., A. Bodenteich, Y.F. Wang, Y.P. Wang, D. Burian, S.W. Clifton, J. Crabtree, A. Freeman, K. Iyer, L. Jian, Y. May, H.J. McLaury, H.Q. Pan, O. Sharan, S. Toth, Z. Wong, G. Zhang, N. Heisterkamp, J. Groffen, and B.A. Roe. 1995. Sequence and analysis of the human ABL gene, the BCR gene, and regions involved in the Philadelphia chromosomal translocation. Genomics 27:67–82.

Diers, B.W., P. Keim, W.R. Fehr, and R.C. Shoemaker. 1992. RFLP analysis of soybean seed protein and oil content. Theor. Appl. Genet. 83:608–612.

Dong, Q., C.J. Lawrence, S.D. Schlueter, M.D. Wilkerson, S. Kurtz, C. Luschbough, and V. Brendel. 2005. Comparative plant genomics resources at PlantGDB. Plant Physiol. 139:610–618.

Force, A., M. Lynch, F.B. Pickett, A. Amores, Y.-L. Yan, and J. Postlethwait. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531–1545.

Foster-Hartnett, D., J. Mudge, D. Danesh, H. Yan, D. Larsen, R. Denny, and N.D. Young. 2002. Comparative genomic analysis of sequences sampled from a small region on soybean molecular linkage group 'G'. Genome 45:634–645.

Frazer, K.A., L. Pachter, A. Poliakov, E.M. Rubin, and I. Dubchak. 2004. VISTA: Computational tools for comparative genomics. Nucleic Acids Res. 32:W273–W279.

Graham, M.A., L.F. Marek, and R.C. Shoemaker. 2002. Organization, expression and evolution of a disease resistance gene cluster in soybean. Genetics 162:1961–1977.

Grandbastien, M.A., S. Berry-Lowe, B.W. Shirley, and R. Meagher. 1986. Two soybean ribulose-1,5-bisphosphate carboxylase small subunit genes share extensive homology even in distant flanking sequences. Plant Mol. Biol. 7:451–465.

Grundy, S.M. 1986. Comparison of monounsaturated fatty acids and carbohydrates for lowering plasma cholesterol in man. N. Engl. J. Med. 314:745–748.

Haberer, G., S. Young, A.K. Bharti, H. Gundlach, C. Raymond, G. Fuks, E. Butler, R.A. Wing, S. Rounsley, B. Birren, C. Nusbaum, K.F.X. Mayer, and J. Messing. 2005. Structure and architecture of the maize genome. Plant Physiol. 139:1612–1624.

Hadley, H.H., and T. Hymowitz. 1973. Speciation and cytogenetics. p. 96–116. *In* B.E. Caldwell (ed.) Soybeans: Improvement, production, and uses. Agron. Monogr. 16. ASA, Madison, WI.

Heppard, E.P., A.J. Kinney, K.L. Stecca, and G.-H. Miao. 1996. Developmental and growth temperature regulation of two different microsomal ω-6 desaturase genes in soybeans. Plant Physiol. 110:311–319.

Hightower, R., and R. Meagher. 1985. Divergence and differential expression of soybean actin genes. EMBO J. 4:1–8.

Lackey, J.A. 1980. Chromosome numbers in the Phaseoleae (Fabaceae:Faboideae) and their relation to taxonomy. Am. J. Bot. 67:595–602.

Lander, E., P. Green, J. Abrahamson, A. Barlow, M. Daly, S. Lincoln, and L. Newberg. 1987. MAPMAKER: An interactive computer package for transcripting primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181.

Lee, J.M., A. Bush, J.E. Specht, and R.C. Shoemaker. 1999. Mapping duplicate genes in soybean. Genome 42:829–836.

Lee, J.M., D. Grant, C.E. Vallejos, and R.C. Shoemaker. 2001. Genome organization in dicots. II. *Arabidopsis* as a 'bridging species' to resolve genome evolution events among legumes. Theor. Appl. Genet. 103:765–773.

Lee, J.S., and D.P.S. Verma. 1984. Chromosomal arrangement of leghemoglobin genes in soybean. Nucleic Acids Res. 11:5541–5553.

Lin, J.-Y., B.H. Jacobus, P. SanMiguel, J.G. Walling, Y. Yuan, J. Doyle, R.C. Shoemaker, N.D. Young, and S.A. Jackson. 2005. Pericentromeric regions of soybean (*Glycine max* L. Merr.) chromosomes consist of retroelements and tandemly repeated DNA and are structurally and evolutionarily labile. Genetics 170:1221–1230.

Lockton, S., and B.S. Gaut. 2005. Plant conserved non-coding sequences and paralogue evolution. Trends Genet. 21:60–65.

Lukashin, A., and M. Borodovsky. 1998. GeneMark.hmm: New solutions for gene finding. Nucleic Acids Res. 26:1107–1115.

Marek, L.F., J. Mudge, L. Darnielle, D. Grant, N. Hanson, M. Paz, Y. Huihuang, R. Denny, K. Larson, D. Foster-Hartnett, A. Cooper, D. Danesh, D. Larsen, T. Schmidt, R. Staggs, J.A. Crow, E. Retzel, N.D. Young, and R.C. Shoemaker. 2001. Soybean genomic survey: BAC-end sequences near RFLP and SSR markers. Genome 44:572–581.

Marek, L.F., and R.C. Shoemaker. 1997. BAC contig development by fingerprint analysis in soybean. Genome 40:420–427.

Masterson, J. 1994. Stomatal size in fossil plants: Evidence for polyploidy in majority of angiosperms. Science (Washington, DC) 264:421–424.

Mattson, F.H., and S.M. Grundy. 1985. Comparison of effects of dietary saturated, monounsaturated, and polyunsaturated fatty acids on plasma lipids and lipoproteins in man. J. Lipid Res. 26:194–202.

Mattson, J.W., C. Sun, and W.W. Koo. 2004. Analysis of the world oil crops market. Rep. No. 529. Dep. of Agribusiness and Applied Economics, North Dakota State Univ., Fargo.

McCarthy, E.M., and J.F. McDonald. 2003. LTR_STRUC: A novel search and identification program for LTR retrotransposons. Bioinformatics 19:362–367.

Messing, J., A.K. Bharti, W.M. Karlowski, H. Gundlach, H.R. Kim, Y. Yu, F. Wei, G. Fuks, C.A. Soderlund, K.F.X. Mayer, and R.A. Wing. 2004. Sequence composition and genome organization of maize. Proc. Natl. Acad. Sci. USA 101:14349–14354.

Miquel, M.F., and J.A. Browse. 1994. High-oleate oilseeds fail to develop at low temperature. Plant Physiol. 106:421–427.

Mudge, J., S.B. Cannon, P. Kalo, G.E.D. Oldroyd, B.A. Roe, C.D. Town, and N.D. Young. 2005. Highly syntenic regions in the genomes of soybean, *Medicago truncatula* and *Arabidopsis thaliana*. BMC Plant Biol. 5:15.

Neidleman, S.L. 1987. Effects of temperature on lipid unsaturation. Biotechnol. Genet. Eng. Rev. 5:245–268.

Nielsen, N.C., C. Dickinson, T.J. Cho, V.H. Thanh, B.J. Scallon, R.L. Fischer, T.L. Sims, G.N. Drews, and R.B. Goldberg. 1989. Characterization of the glycinin gene family in soybean. Plant Cell 1:313–328.

Okuley, J., J. Lightner, K. Feldmann, N. Yadav, and J. Browse. 1994. The Arabidopsis FAD2 gene encodes the enzyme that is essential for polyunsaturated lipid synthesis. Plant Cell 6:147–158.

Pan, H.Q., Y.P. Wang, S.L. Chissoe, A. Bodenteich, Z. Wang, K. Iyer, S.W. Chifton, J.S. Crabtree, and B.A. Roe. 1994. The complete nucleotide sequence of the SacBII domain of the P1 pAD10-SacBII Cloning vetor and three cosmid cloning vectors: PTCF, svPHEP, and LAWRIST16. GATA 11:181–186.

Roe, B.A. 2004. Shotgun library transcription for DNA sequencing. p. 171–187. *In* X. Zhao and M. Stodolsky (ed.) Methods in molecular biology. Vol. 255. Bacterial artificial chromosomes, Vol. 1. Library transcription, physical mapping, and sequencing. Human Press, Totowa, NJ.

Roe, B., J. Crabtree, and A. Khan. 1996. *In* D. Rickwood (ed.) DNA isolation and sequencing: Essential techniques series. J. Wiley & Sons, NY.

SanMiguel, P., and J.L. Bennetzen. 1998. Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. Ann. Bot. (Lond.) 82:37–44.

Schlueter, J.A., P. Dixon, C. Granger, D. Grant, L. Clark, J.J. Doyle, and R.C. Shoemaker. 2004. Mining EST databases to resolve evolutionary events in major crop species. Genome 47:868–876.

Schlueter, J.A., B.E. Scheffler, and R.C. Shoemaker. 2006. Sequence conservation of homeologous BACs and expression of homeologous genes in soybean (*Glycine max* L. Merr). Genetics (in press) doi:10.1534/genetics.105.055020.

Schlueter, S.D., Q. Dong, and V. Brendel. 2003. GeneSeqer@PlantGDB: Gene structure predication in plant genomes. Nucleic Acids Res. 31:3597–3600.

Schlueter, S.D., M.D. Wilkerson, E. Huala, S.Y. Rhee, and V. Brendel. 2005. Community-based gene structure annotation for the *Arabidopsis thaliana* genome. Trends Plant Sci. 10:9–14.

Shoemaker, R., K. Polzin, J. Labate, J. Specht, E.C. Brummer, T. Olson, N. Young, V. Concibido, J. Wilcox, J. Tamulonis, G. Kochert, and H.R. Boerma. 1996. Genome duplication in soybean (*Glycine* subgenus *soja*). Genetics 144:329–338.

Song, Q.J., L.F. Marek, R.C. Shoemaker, K.G. Lark, V.C. Concibido, X. Delannay, J.E. Specht, and P.B. Cregan. 2004. A new integrated genetic linkage map of the soybean. Theor. Appl. Genet. 109:122–128.

Su, Z., J. Wang, J. Yu, X. Huang, and X. Gu. 2006. Evolution of alternative splicing after gene duplication. Genome Res. 16:182–189.

Tang, G.-Q., W.P. Novitzky, H.C. Griffing, S.C. Huber, and R.E. Dewey. 2005. Oleate desaturase enzymes of soybean: Evidence of regulation through differential stability and phosphorylation. Plant J. 44:433–446.

Thomas, J.M.G., K.J. Boote, L.H. Allen, M. Gallo-Meagher, and J.M. David. 2003. Seed physiology and metabolism: Elevated temperature and carbon dioxide effects on soybean seed composition and transcript abundance. Crop Sci. 43:1548–1577.

Thompson, G.A. 1993. Response of lipid metabolism to developmental change and environmental perturbation. p. 591–619. *In* T.S. Moore (ed.) Lipid metabolism in plant. CRC Press, Boca Raton, FL.

Walling, J.G., R. Shoemaker, N. Young, J. Mudge, and S. Jackson. 2006. Chromosome-level homeology in paleopolyploid soybean (*Glycine max*) revealed through integration of genetic and chromosome maps. Genetics 172:1893–1900.

Warner, K., P. Orr, L. Parrot, and M. Glynn. 1994. Effects of frying oil composition on potato chip stability. J. Am. Oil Chem. Soc. 71:1117–1121.

Wilkerson, M.D., S.D. Schlueter, and V. Brendel. 2006. yrGATE: A web-based gene-structure annotation tool for the identification and dissemination of eukaryotic genes. Genome Biol. 7:R58.

Wolfe, K.H. 2001. Yesterday's polyploids and the mystery of diploidization. Nat. Rev. Genet. 2:333–341.

Yan, H.H., J. Mudge, D.J. Kim, D. Larsen, R.C. Shoemaker, D.R. Cook, and N.D. Young. 2003. Estimates on conserved microsynteny among the genomes of *Glycine max, Medicago truncatula* and *Arabidopsis thaliana*. Theor. Appl. Genet. 106:1256–1265.

Yan, H.H., J. Mudge, D.J. Kim, R.C. Shoemaker, D.R. Cook, and N.D. Young. 2004. Comparative physical mapping reveals features of microsynteny between *Glycine max, Medicago truncatula*, and *Arabidopsis thaliana*. Genome 47:141–155.

Yang, Z. 1997. PAML: A program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 15:555–556.

Young, N.D., J. Mudge, and Y.N. Ellis. 2003. Legume genomes: More than peas in a pod. Curr. Opin. Plant Biol. 6:199–204.